

Serial No.: 09/468,785
Atty. Docket No.: P64616US0

REMARKS

The Office Action mailed November 7, 2003, has been carefully reviewed and by this Amendment, Applicants have canceled claims 5-15, 20-30 and 35-45, amended claims 1-4, 16-19 and 31-34, and added claims 46-60. Claims 1-4, 16-19, 31-34 and 46-60 are pending in the application. Claims 1, 2, 16, 17, 31 and 32 are the independent claims. In view of the above amendments and the following remarks, favorable reconsideration in this application is respectfully requested.

The Examiner rejected claims 1, 16 and 31 under 35 U.S.C. 102(e) as being anticipated by U.S. Patent No. 6,289,342 to Lawrence et al. ("Lawrence"), and rejected claims 2-4, 17-19 and 32-34 under 35 U.S.C. 103(a) as being unpatentable over Lawrence in view of U.S. Patent No. 5,638,543 to Pedersen et al. ("Pedersen").

As set forth in the amended independent claims, the present invention is directed to a method, apparatus and computer-readable code for identifying text, taken from a set of citing documents and appearing near a citing instance, that indicates the reasons for citing (RFC) the particular document represented by the citing instance.

According to the method, the contexts of the citing instances in the respective citing documents are obtained, with each context including at least one text unit that includes the citing instance and at least one text unit that is near the citing instance. The content of the contexts is then analyzed in order to determine a content score for each text unit. More specifically, the content analysis includes identifying those text unit content words that are common to at least two of the citing documents' contexts or to at least one citing document's

Serial No.: 09/468,785
Atty. Docket No.: P64616US0

context and the cited document itself. In other words, in the former instance, there is a comparison between the context of a citing instance in one citing document and the context of another citing instance appearing in a different citing document; in the latter, there is a comparison between the context of a citing instance in one citing document and the text of the document actually being cited. In either case, the texts being compared come from two different documents. Based on the comparison process, a content score is calculated for each text unit and at least one text unit is selected from the citing documents' contexts as constituting the RFC on the basis of the scoring. This is not shown or suggested by the prior art of Lawrence, whether taken alone or in combination with Pedersen.

Lawrence discloses an autonomous citation indexing system directed to identifying published article citations with varying formats and correlating those which, though differing in syntax, are making reference to the same paper. While the Lawrence system extracts portions of the text surrounding the citation so as to enable a searcher to make a better determination of the nature of the publication citing the article, contrary to the Examiner's conclusion, Lawrence does not teach or suggest the step of analyzing the content of the contexts which, as clarified through the amendments to claims 1, 2, 16, 17, 31 and 32 set forth herein, includes comparing portions of text taken from two documents to identify content words that are common to both documents.

As a cite indexing system, the purpose of Lawrence is to identify various article citations within a referencing article so that it may be determined where and how frequently an article is cited within the greater body of literature. However, there is nothing in Lawrence that

Serial No.: 09/468,785
Atty. Docket No.: P64616US0

speaks to identifying the specific reason for which the article was cited through comparison with one or more citing instances appearing in other citing documents.

The Examiner has identified Figure 6 of Lawrence as showing the selection of sentences that constitute the RFC, but Applicants respectfully request reconsideration of what is fairly taught in Lawrence with respect to what is claimed by the present invention. As shown in Figure 6, Lawrence teaches the extraction of an article cite, along with the immediately surrounding text in its original format, for display in order to allow a researcher to scan such text to see the context of the article citation. However, there is nothing in this side-by-side display of two text sets that teaches or suggests either an automated comparison of the words within the two text sets to identify commonality, or the automated scoring of text units (which may be sentences) on the basis of common content words shared by the text sets in order to identify the RFC, as claimed by the present invention.

With respect to the comparison of two texts, the Examiner has relied upon column 8, lines 12-16, as teaching the generation of a content word list based on words that are included in the contexts of at least two of the citing documents, and upon column 9, lines 28-31, as teaching the generation of a content word list based on words that are included in the context of at least one of the citing documents and the cited document. Applicants are unsure of the basis upon which the Examiner finds this referenced text to teach the subject matter claimed by the present invention and request clarification.

Specifically, from Applicants' reading, column 8, lines 12-16, of Lawrence summarizes the document information which is stored in a "new database" when a CiteSeer

Serial No.: 09/468,785
Atty. Docket No.: P64616US0

agent is created. Assuming the Examiner considers the “new database” to be analogous to the content word list, there is nothing in column 8, lines 12-16 that requires that words entered into this “new database” be included in the contexts of at least two citing documents or in the context of one citing document and the cited document. Nor is there anything in column 9, lines 28-31, that requires or suggests such document cross-comparison as a predicate for inclusion in a subsequent listing. Giving full consideration to what is being claimed, the present invention goes well beyond what is fairly taught by Lawrence and could only be said to be suggested by Lawrence on the basis of Applicants’ own teaching, which cannot be relied upon.

Pedersen also fails to provide the necessary teaching, being directed to a method and apparatus for summarizing documents on an automated basis through the scoring and classifying of text regions *within an individual document* in order to generate an abstract thereof. The region selection and scoring set forth in column 3, lines 16-18, cited by the Examiner, speaks to this abstract generation step. What is completely absent from Pedersen is anything that would suggest comparing text units taken from *two different* documents. Rather, Pedersen’s abstract generating method and apparatus is necessarily limited to the single document at issue in that comparison with the text of other documents would be of no value or purpose give the intended function of Pedersen.

Applicants also object to the basis of the Examiner’s finding of motivation to combine Lawrence and Pedersen, expressed in connection with the rejection of claims 3, 4, 18, 19, 33 and 34, as explicitly relying upon Applicants’ own disclosure which is not permitted. Again, Lawrence is concerned with recognizing article citations having varying formats as

Serial No.: 09/468,785
Atty. Docket No.: P64616US0

representing the same article; there is no need to analyze the words contained in the surrounding text and to do so would not enhance the article citation recognition function. Pedersen is limited to automated abstract generation based on the text within the target document and effectively teaches against any comparison of text from at least two documents as being irrelevant and even counter-productive to the expressed purpose. For at least the foregoing reasons, claims 1, 2, 16, 17, 31 and 32 are not obvious in view of the prior art but are patentable othereover; thus reconsideration of the teaching of Lawrence and Pedersen with respect to the independent claims, and withdrawal of the rejection thereof, is sincerely requested. Claims 3, 4, 18, 19, 33, and 34 are in condition for allowance as claims properly dependent on an allowable base claim and for the subject matter contained therein.

New claims 46-60 are also in condition for allowance as claims properly dependent on an allowable base claim and for the subject matter contained therein. Specifically, the prior art does not teach or suggest the specific manner in which the content word list is generated by associating paragraphs from the documents, eliminating noise words, determining common words found in at least two paragraphs and tallying a frequency count for such words which are linked with the common words to form the content word list, as set forth in claims 46, 51 and 56, nor the stemming of the content words in conjunction with the remaining steps as set forth in claims 47, 52 and 57.

Nor does the prior art disclose the specific manner by which the content scores are calculated, namely by calculating respective initial content scores (ICS) for the sentences in the citing documents based on the content words in the sentences, calculating respective distances of

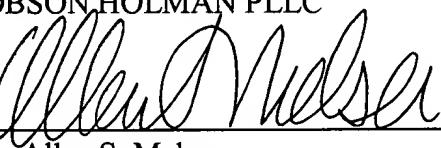
Serial No.: 09/468,785
Atty. Docket No.: P64616US0

the sentences in the citing documents from respective citing instances of the cited document, and calculating respective content scores (CS) for the sentences in the citing documents, based on at least the ICS and the distances, as set forth in claims 48, 53, and 58. The additional limitations set forth in claims 49, 54 and 59, of normalizing the ICS to form normalized initial content scores (NICS) for use by the CS calculation step, the normalizing process taking into account numbers of words in the respective sentences and a largest frequency count in the content word list, as well as the process of modifying the distances to form respective modified absolute distances using criteria that include whether a sentence is in a same paragraph with the citing instance or is located after the citing instance (claims 50, 55 and 60), are also well beyond the scope of what is shown in the prior art. Favorable consideration and allowance of claims 46-60 is requested.

With this Amendment, the application is in condition for allowance. Should the Examiner have any questions or comments, the Examiner is cordially invited to telephone the undersigned attorney so that the present application can receive an early Notice of Allowance.

Respectfully submitted,

JACOBSON HOLMAN PLLC

By 
Allen S. Melser
Reg. No. 27,215

400 Seventh Street, N.W.
Washington, D.C. 20004
Telephone: (202) 638-6666
Date: February 9, 2004

R:\SBAILEY\2-04\P64616US.amd